A diagrammatic approach to study the information transfer in weakly non-linear channels

E. KORUTCHEVA

Departamento de Física Fundamental
Universidad Nacional de Educación a Distancia
c/Senda del Rey No 9 - 28080 Madrid, Spain *

VALERIA DEL PRETE

Institute for theoretical physics
University of Leuven
Celestijnenlaan 200 D, B-3001 Leuven, Belgium

*electronic address: elka@fisfun.uned.es; Permanent address: G. Nadjakov Institute of Solid State Physics, Bulgarian Academy of Sciences, 1784 Sofia, Bulgaria

Abstract

In a recent work we have introduced a novel approach to study the effect of weak non-linearities in the transfer function on the information transmitted by an analogue channel, by means of a perturbative diagrammatic expansion. We extend here the analysis to all orders in perturbation theory, which allows us to release any constraint concerning the magnitude of the expansion parameter and to establish the rules to calculate easily the contribution at any order. As an example we explicitly compute the information up to the second order in the non-linearity, in presence of random gaussian connectivities and in the limit when the output noise is not small. We analyze the first and second order contributions to the mutual information as a function of the non-linearity and of the number of output units. We believe that an extensive application of our method via the analysis of the different contributions at distinct orders might be able to fill a gap between well known analytical results obtained for linear channels and the non trivial treatments which are required to study highly non-linear channels.

1 Introduction

Several recent investigations have explored the efficiency of two-layer networks in transmitting information, given that the distribution of the input layer and the input-output transformation are known [1, 2, 3, 4, 5, 6, 7, 8, 9]. Some of these works have been inspired by processes involving populations of real neurons in the brain [1, 2, 3]. There are experimental evidences that real neurons operate a highly non-linear transformation of the inputs, whose main features can be well captured by a threshold-linear function. Moreover this type of transformation allows an easier analytical treatment, at least under the assumption of replica symmetry [1, 2, 3]. A sigmoidal transfer function has been proposed in other works and in the more general context of neural networks [10, 11]. Yet this choice makes the analytical treatment much more difficult.

In a more theoretical framework, recent studies have explored the communication properties of linear and binary channels, where the transfer function is either purely linear [4, 5], or highly non-linear [6, 7, 8]. The analytical solution in case of pure linearity is straightforward, while in presence of a step transfer function one must resort to some approximations like replica symmetry, or restrict oneself to some particular regions in the parameter space.

No extensive study has been performed yet, trying to bridge these two limit cases, pure linearity or strong non-linearity in the transfer function, with respect to the impact on the information content, i.e. on the channel efficiency.

In a very recent study [9], the contribution of small non-linearities to the mutual information has been evaluated in case of a gaussian noisy channel, introducing a novel approach by means of a perturbative expansion and providing an elegant interpretation in terms of feynman diagrams [12]. An analytical expression for the mutual information has been obtained at first order in the non-linearity.

Here we extend the previous analyses to all orders in perturbation theory, deriving both the analytical expansion and the diagrammatic formalism necessary to evaluate the contribution to the mutual information at each order. Then we apply our method and quantify the first and second order contributions to the mutual information in the case of random gaussian connectivities and in the limit of large output noise.

Even though not motivated by any particular hardware or biological application, our study is an attempt to fill a theoretical gap between purely linear and strongly non-linear channels, at least in the case of gaussian units.

We believe that an extensive application of our expansion will allow to examine the impact of non-linearities on the information transmission and its modulation with the noise and the other parameters in the model. This will be the object of future investigations.

2 The model

The network model is analogous to the one used in [9].

The distribution of the N continuous inputs $\boldsymbol{x}=\{x_1...x_N\}$ is gaussian with correlation matrix C, and each input signal is corrupted by uncorrelated gaussian noise $\boldsymbol{\nu}=\{\nu_1..\nu_N\}$, as in the following:

$$\langle x_i \rangle = 0, \tag{1}$$
$$\langle x_i x_j \rangle = [C]_{ij}, \quad \forall i, j \in 1, 2, ..N. \tag{2}$$

$$\langle \nu_i \rangle = 0, \tag{3}$$
$$\langle \nu_i \nu_j \rangle = b_0\, \delta_{ij}, \quad \forall i, j \in 1, 2, ..N. \tag{4}$$

Each output unit performs a linear summation of the noisy inputs $\boldsymbol{x} + \boldsymbol{\nu}$ via the matrix of the connectivities $J = \{J_{ij}\}$; the result is transformed via a non-linear transfer function G:

$$\boldsymbol{V} = G(J(\boldsymbol{x} + \boldsymbol{\nu})) + \boldsymbol{z}, \tag{5}$$

where $\boldsymbol{z} = \{z_1 .. z_N\}$ is uncorrelated gaussian noise affecting the output, according to the following distribution:

$$\langle z_i \rangle = 0, \tag{6}$$
$$\langle z_i z_j \rangle = b\, \delta_{ij}, \forall i, j \in 1, 2, ..N. \tag{7}$$

In analogy to the case examined in [9], we consider a sigmoidal shaped transfer function assuming that the non-linearity is small, so that a Taylor expansion can be performed:

$$G(x) = th(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}. \tag{8}$$

For small values of x one can stop at the first non-linear term, the cubic one. To illustrate the method we will first focus on a cubic non-linearity for the sake of simplicity. Nonetheless, as it will be clear in the following, out method is equally applicable to any term in the expansion of the hyperbolic tangent, finally allowing to reconstruct the whole series.

In a cubic approximation eq.(5) can be rewritten as:

$$\boldsymbol{V} \simeq \boldsymbol{h} + \boldsymbol{g}(\boldsymbol{h}) + \boldsymbol{z}, \tag{9}$$

with:

$$\boldsymbol{g}(\boldsymbol{h}) = g_0 \boldsymbol{h}^3,$$
$$\boldsymbol{h} = J(\boldsymbol{x} + \boldsymbol{\nu}). \tag{10}$$

3 A perturbative approach to the mutual information

In analogy to what is reported in [9] we express the mutual information between input and output as the difference between the *output entropy* and the *equivocation* between input and output[1]:

---

[1]Using the natural logarithm we implicitly measure the information in natural numbers. Conversion to bits is easily obtained dividing the mutual information by $\ln 2$:

$$I = \int d\boldsymbol{x} \int d\boldsymbol{V} P(\boldsymbol{x}, \boldsymbol{V}) \ln \frac{P(\boldsymbol{x}, \boldsymbol{V})}{P(\boldsymbol{x})P(\boldsymbol{V})} = H(\boldsymbol{V}) - \langle H(\boldsymbol{V}|\boldsymbol{x}) \rangle_x; \tag{11}$$

$$H(\boldsymbol{V}) = - \int d\boldsymbol{V} P(\boldsymbol{V}) \ln P(\boldsymbol{V}) \tag{12}$$

$$\langle H(\boldsymbol{V}|\boldsymbol{x}) \rangle_x = - \int d\boldsymbol{x} \int d\boldsymbol{V} P(\boldsymbol{x}) P(\boldsymbol{V}|\boldsymbol{x}) \ln P(\boldsymbol{V}|\boldsymbol{x}). \tag{13}$$

The probabilities $P(\boldsymbol{V})$, $P(\boldsymbol{x})$, $P(\boldsymbol{V}|\boldsymbol{x})$ are fixed by eqs.(2), (4),(9); in particular explicit expressions for both $P(\boldsymbol{V})$ and $P(\boldsymbol{V}|\boldsymbol{x})$ can be obtained explicitating the dependence of $\boldsymbol{V}$ on $\boldsymbol{h}$. One has:

$$P(\boldsymbol{V}) = \int d\boldsymbol{h} P(\boldsymbol{h}) P(\boldsymbol{V}|\boldsymbol{h}); \tag{14}$$

$$P(\boldsymbol{V}|\boldsymbol{x}) = \int d\boldsymbol{h} P(\boldsymbol{h}|\boldsymbol{x}) P(\boldsymbol{V}|\boldsymbol{h}); \tag{15}$$

$$P(\boldsymbol{V}|\boldsymbol{h}) = \frac{1}{\sqrt{2\pi b^N}} \cdot e^{-(\mathbf{V}-\mathbf{h}-\mathbf{g}(\mathbf{h}))^2/2b}. \tag{16}$$

where $\boldsymbol{h} = J(\boldsymbol{x} + \boldsymbol{\nu})$. Given the relationships (2), (4) it is trivial to derive that both $P(\boldsymbol{h})$ and $P(\boldsymbol{h}|\boldsymbol{x}) = P(\boldsymbol{h} - J\boldsymbol{x})$ are gaussian distributed:

$$P(\boldsymbol{h}) = \frac{e^{-\frac{1}{2}\mathbf{h}(\mathbf{A})^{-1}\mathbf{h}}}{\sqrt{(2\pi)^N detA}}; \quad A = JCJ^T + b_0 JJ^T, \tag{17}$$

$$P(\boldsymbol{h}|\boldsymbol{x}) = \frac{e^{-\frac{1}{2}(\mathbf{h}-\mathbf{Jx})(B)^{-1}(\mathbf{h}-\mathbf{Jx})}}{\sqrt{(2\pi)^N detB}}; \quad B = b_0 JJ^T. \tag{18}$$

It is just the presence of only gaussian averages that will finally allow us to express the mutual information as a series of Feynman diagrams by means of the Wick theorem.

Further details can be found in [9].

3.1   Perturbative expansion for the output entropy

Let us consider eq.(12). When the non-linear term $\boldsymbol{g}(\boldsymbol{h})$ is non zero, integration on $\boldsymbol{h}$ cannot be performed without resorting to some approximation. In [9] it has been shown that an expansion up to first order in $\boldsymbol{g}(\boldsymbol{h})$ allows to perform the integration and derive an analytical expression for the mutual information. Our purpose here is to extend the previous analysis, considering all terms in the expansion, of whatever order in $\boldsymbol{g}(\boldsymbol{h})$.

Let us consider the following equalities:

$$P(\boldsymbol{V}) = \int d\boldsymbol{h} P(\boldsymbol{h}) P_0(\boldsymbol{V}|\boldsymbol{h}) e^{\Delta/b} = \int d\boldsymbol{h} P(\boldsymbol{h}) P_0(\boldsymbol{V}|\boldsymbol{h}) \left[\sum_{l=0}^{\infty} \frac{\Delta^l}{l!\, b^l} \right], \tag{19}$$

where

$$P_0(\boldsymbol{V}|\boldsymbol{h}) = \frac{1}{\sqrt{2\pi b^N}} \cdot e^{-(\mathbf{V}-\mathbf{h})^2/2b}, \tag{20}$$

$$\Delta = \Delta(\boldsymbol{V}, \boldsymbol{h}) = (\boldsymbol{V} - \boldsymbol{h}) \cdot \boldsymbol{g}(\boldsymbol{h}) - \frac{\boldsymbol{g}(\boldsymbol{h})^2}{2} \tag{21}$$

and we have performed the perturbation expansion in terms of Δ, keeping in mind that an explicit expression in terms of powers of $\boldsymbol{g}$ can be extracted a posteriori.

Inserting the expansion in powers of Δ, eq.(19), in eq.(12) it can be shown that the output entropy can be expressed as follows:

$$\begin{aligned}
H(\boldsymbol{V}) &= H_0(\boldsymbol{V}) - \int d\boldsymbol{V} \left(\sum_{l=1}^{\infty} \frac{\left\langle \Delta^l \right\rangle_h}{l!\, b^l} \right) \ln P_0(\boldsymbol{V}) \\
&+ \int d\boldsymbol{V} P_0(\boldsymbol{V}) \sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n(n-1)} \left(\sum_{l=1}^{\infty} \frac{1}{P_0(\boldsymbol{V})} \frac{\left\langle \Delta^l \right\rangle_h}{l!\, b^l} \right)^n,
\end{aligned} \tag{22}$$

where we have used the notation:

$$\left\langle \Delta^l \right\rangle_h = \int d\boldsymbol{h} P(\boldsymbol{h}) P_0(\boldsymbol{V}|\boldsymbol{h}) \Delta^l = \int d\boldsymbol{h} P(\boldsymbol{h}) P_0(\boldsymbol{V}|\boldsymbol{h}) \left[(\boldsymbol{V} - \boldsymbol{h}) \cdot \boldsymbol{g}(\boldsymbol{h}) - \frac{\boldsymbol{g}(\boldsymbol{h})^2}{2} \right]^l. \tag{23}$$

$H_0(\boldsymbol{V})$ is the output entropy when $\boldsymbol{g} = 0$:

$$H_0(\boldsymbol{V}) = \int d\boldsymbol{V} P_0(\boldsymbol{V}) \ln P_0(\boldsymbol{V}). \tag{24}$$

From eqs.(20),(17) it is easy to derive that:

$$P_0(\boldsymbol{V}) = \int d\boldsymbol{h} P(\boldsymbol{h}) P_0(\boldsymbol{V}|\boldsymbol{h}) \frac{e^{-\frac{1}{2}\mathbf{V}(\mathbf{A}+\mathbf{bI})^{-1}\mathbf{V}}}{\sqrt{(2\pi)^N det(A + bI)}}, \tag{25}$$

which allows to derive an explicit expression for $H_0(\boldsymbol{V})$:

$$H_0(\boldsymbol{V}) = \frac{N}{2} \left[1 + \ln 2\pi \right] + \frac{1}{2} det(A + bI). \tag{26}$$

In the limit when g_0 is very small and $g_0 \ll b$ one can stop at the first order in Δ, neglecting the second order term in $\boldsymbol{g}$. The expression for the output entropy becomes:

$$H(\boldsymbol{V}) \simeq H_0(\boldsymbol{V}) - \frac{1}{b} \int d\boldsymbol{h} \int d\boldsymbol{V} P(\boldsymbol{h}) P_0(\boldsymbol{V}|\boldsymbol{h}) \boldsymbol{g}(\boldsymbol{h})(\boldsymbol{V} - \boldsymbol{h}) \ln P_0(\boldsymbol{V}). \tag{27}$$

In this simpler case, as it has been shown in [9], integration is straightforward, leading to an explicit final expression for the output entropy, as well as for the mutual information. Here we will show that it is possible to generalize our approach to every order in perturbation theory, establishing the basic rules to identify each integral with a diagram, so that the analytical evaluation is reduced to a combinatorial problem via the use of Wick theorem [12].

Let us reconsider eq.(23). A change of variables allows to reduce the integration on $\boldsymbol{h}$ to a gaussian:

$$\boldsymbol{h} \longrightarrow \boldsymbol{\gamma} + \frac{1}{b}D\boldsymbol{V};$$

$$\left\langle \Delta^l \right\rangle_h \longrightarrow P_0(\boldsymbol{V}) \int d\boldsymbol{\gamma} P(\boldsymbol{\gamma}) \left[g_0 \left(\boldsymbol{V} - \boldsymbol{\gamma} - \frac{1}{b}D\boldsymbol{V} \right) \left(\boldsymbol{\gamma} + \frac{1}{b}D\boldsymbol{V} \right)^3 - \frac{g_0^2}{2} \left(\boldsymbol{\gamma} + \frac{1}{b}D\boldsymbol{V} \right)^6 \right]^l \quad (28)$$

where

$$P(\boldsymbol{\gamma}) = \frac{e^{-\frac{1}{2}\boldsymbol{\gamma} D^{-1} \boldsymbol{\gamma}}}{\sqrt{(2\pi)^N det D}}; \quad D^{-1} = A^{-1} + b^{-1}I. \quad (29)$$

Inserting this expression in eq.(22) one obtains:

$$H(\boldsymbol{V}) = H_0(\boldsymbol{V}) +$$

$$\frac{1}{2} \left\langle \left\langle \sum_{l=1}^{\infty} \frac{g_0^l}{l!\, b^l} \left[\left(\boldsymbol{V} - \boldsymbol{\gamma} - \frac{1}{b}D\boldsymbol{V} \right) \left(\boldsymbol{\gamma} + \frac{1}{b}D\boldsymbol{V} \right)^3 - \frac{g_0}{2} \left(\boldsymbol{\gamma} + \frac{1}{b}D\boldsymbol{V} \right)^6 \right]^l \right\rangle_\gamma \left(\boldsymbol{V} \left[A + bI \right]^{-1} \boldsymbol{V} \right) \right\rangle_V$$

$$+ \left\langle \sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n(n-1)} \left\langle \sum_{l=1}^{\infty} \frac{g_0^l}{l\, b^l} \left[\left(\boldsymbol{V} - \boldsymbol{\gamma} - \frac{1}{b}D\boldsymbol{V} \right) \left(\boldsymbol{\gamma} + \frac{1}{b}D\boldsymbol{V} \right)^3 - \frac{g_0}{2} \left(\boldsymbol{\gamma} + \frac{1}{b}D\boldsymbol{V} \right)^6 \right]^l \right\rangle_\gamma^n \right\rangle_V, \quad (30)$$

where

$$\langle F(\boldsymbol{\gamma}, \boldsymbol{V}) \rangle_\gamma = \int d\boldsymbol{\gamma} P(\boldsymbol{\gamma}) F(\boldsymbol{\gamma}, \boldsymbol{V}),$$

$$\langle F(\boldsymbol{\gamma}, \boldsymbol{V}) \rangle_V = \int d\boldsymbol{V} P_0(\boldsymbol{V}) F(\boldsymbol{\gamma}, \boldsymbol{V}). \quad (31)$$

$P(\boldsymbol{\gamma}), P_0(\boldsymbol{V})$ are given respectively by eqs.(29), (25) and we have explicitly put $\boldsymbol{g}(\boldsymbol{h}) = g_0 \boldsymbol{h}^3$.

3.2 Perturbative expansion for the equivocation

A very similar expression can be obtained for the equivocation, by means of the same perturbative approach.

Let us reconsider eqs.(13),(16). Since $P(\boldsymbol{V}|\boldsymbol{x})$ can be expressed as in integral on the distribution $P(\boldsymbol{V}|\boldsymbol{h})$, we can use the same perturbative expansion already introduced for $P(\boldsymbol{V})$ in eq.(19):

$$P(\boldsymbol{V}|\boldsymbol{x}) = \int d\boldsymbol{h} P(\boldsymbol{h}|\boldsymbol{x}) P_0(\boldsymbol{V}|\boldsymbol{h}) e^{\Delta/b} = \int d\boldsymbol{h} P(\boldsymbol{h}|\boldsymbol{x}) P_0(\boldsymbol{V}|\boldsymbol{h}) \left[\sum_{l=0}^{\infty} \frac{\Delta^l}{l!\, b^l} \right], \quad (32)$$

where $P_0(\boldsymbol{V}|\boldsymbol{h})$ and Δ are given in eqs.(20),(21) and $P(\boldsymbol{h}|\boldsymbol{x})$ is given in eq.(18).

Introducing eq.(32) in the expression for the equivocation one obtains:

$$\langle H(\boldsymbol{V}|\boldsymbol{x}) \rangle_x = \langle H_0(\boldsymbol{V}|\boldsymbol{x}) \rangle_x - \int d\boldsymbol{x} P(\boldsymbol{x}) \int d\boldsymbol{V} \left(\sum_{l=1}^{\infty} \frac{\left\langle \Delta^l \right\rangle_{h|x}}{l!\, b^l} \right) \ln P_0 \left(\boldsymbol{V}|\boldsymbol{x} \right))$$

$$+ \int d\boldsymbol{x} \int d\boldsymbol{V} P_0(\boldsymbol{V}|\boldsymbol{x}) \sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n(n-1)} \left(\sum_{l=1}^{\infty} \frac{1}{P_0(\boldsymbol{V}|\boldsymbol{x})} \frac{\left\langle \Delta^l \right\rangle_{h|x}}{l!\, b^l} \right)^n, \quad (33)$$

7

where we have used the notation:

$$\left\langle \Delta^l \right\rangle_{h|x} = \int d\boldsymbol{h} P(\boldsymbol{h}|\boldsymbol{x}) P_0(\boldsymbol{V}|\boldsymbol{h}) \Delta^l = \int d\boldsymbol{h} P(\boldsymbol{h}|\boldsymbol{x}) P_0(\boldsymbol{V}|\boldsymbol{h}) \left[(\boldsymbol{V} - \boldsymbol{h}) \cdot \boldsymbol{g}(\boldsymbol{h}) - \frac{\boldsymbol{g}(\boldsymbol{h})^2}{2} \right]^l \tag{34}$$

and $\langle H_0(\boldsymbol{V}|\boldsymbol{x})\rangle_x$ is the equivocation when $\boldsymbol{g} = 0$:

$$\langle H_0(\boldsymbol{V}|\boldsymbol{x})\rangle_x = \int d\boldsymbol{x} \int d\boldsymbol{V} P(\boldsymbol{x}) P_0(\boldsymbol{V}|\boldsymbol{x}) \ln P_0(\boldsymbol{V}|\boldsymbol{x}), \tag{35}$$

$$P_0(\boldsymbol{V}|\boldsymbol{x}) = \frac{e^{-\frac{1}{2}(\mathbf{V}-\mathbf{Jx})(B+bI)^{-1}(\mathbf{V}-\mathbf{Jx})}}{\sqrt{(2\pi)^N det(B + bI)}}. \tag{36}$$

Then it is easy to derive the final expression for the zeroth order equivocation:

$$\langle H_0(\boldsymbol{V}|\boldsymbol{x})\rangle_x = \frac{N}{2}(1 + \ln 2\pi) + \frac{1}{2} \ln det(B + bI). \tag{37}$$

On the other hand, if one keeps the contributions to the information up to first order in g_0 one obtains:

$$\langle H(\boldsymbol{V}|\boldsymbol{x})\rangle_x \simeq \langle H_0(\boldsymbol{V}|\boldsymbol{x})\rangle_x - \frac{1}{b} \int d\boldsymbol{h} \int d\boldsymbol{x} \int d\boldsymbol{V} P(\boldsymbol{x}) P(\boldsymbol{h}|\boldsymbol{x}) P_0(\boldsymbol{V}|\boldsymbol{h}) \boldsymbol{g}(\boldsymbol{h}) (\boldsymbol{V}-\boldsymbol{h}) \ln P_0(\boldsymbol{V}|\boldsymbol{x}). \tag{38}$$

Eq.(33) can be treated in a very similar way as already done in the case of the output entropy, eq.(22). Since three different integrations are present in eq.(33) diagonalization of the Gaussian distributions requires replacing both $\boldsymbol{h}$ and $\boldsymbol{V}$ in sequence:

$$\boldsymbol{h} \longrightarrow \tilde{\boldsymbol{\gamma}} + \left(\frac{\boldsymbol{V}}{b} + B^{-1}J\boldsymbol{x}\right)\tilde{D}, \quad \tilde{D}^{-1} = B^{-1} + b^{-1}I,$$

$$\boldsymbol{V} \longrightarrow \boldsymbol{u} + J\boldsymbol{x}. \tag{39}$$

By means of these substitutions eq.(33) can be rewritten as follows:

$$\langle H(\boldsymbol{V}|\boldsymbol{x})\rangle_x = \langle H_0(\boldsymbol{V}|\boldsymbol{x})\rangle_x +$$
$$\frac{1}{2}\left\langle\left\langle \sum_{l=1}^{\infty} \frac{g_0^l}{l!\,b^l}\left[\left(\boldsymbol{u}-\tilde{\boldsymbol{\gamma}}-\frac{1}{b}\tilde{D}\boldsymbol{u}\right)\left(\tilde{\boldsymbol{\gamma}}+\frac{1}{b}\tilde{D}\boldsymbol{u}+J\boldsymbol{x}\right)^3 - \frac{g_0}{2}\left(\tilde{\boldsymbol{\gamma}}+\frac{1}{b}\tilde{D}\boldsymbol{u}+J\boldsymbol{x}\right)^6\right]^l\right\rangle_{\tilde{\gamma},x}\left(\boldsymbol{u}\,[B+bI]^{-1}\,\boldsymbol{u}\right)\right\rangle_u$$
$$+\left\langle \sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n(n-1)}\left\langle \sum_{l=1}^{\infty} \frac{g_0^l}{l!\,b^l}\left[\left(\boldsymbol{u}-\tilde{\boldsymbol{\gamma}}-\frac{1}{b}\tilde{D}\boldsymbol{u}\right)\left(\tilde{\boldsymbol{\gamma}}+\frac{1}{b}\tilde{D}\boldsymbol{u}+J\boldsymbol{x}\right)^3 - \frac{g_0}{2}\left(\tilde{\boldsymbol{\gamma}}+\frac{1}{b}\tilde{D}\boldsymbol{u}+J\boldsymbol{x}\right)^6\right]^l\right\rangle_{\tilde{\gamma}}^n\right\rangle_{u,x} \tag{40}$$

where

$$\langle F(\tilde{\boldsymbol{\gamma}}, \boldsymbol{u}, \boldsymbol{x})\rangle_{\tilde{\gamma}} = \int d\tilde{\boldsymbol{\gamma}} P(\tilde{\boldsymbol{\gamma}}) F(\tilde{\boldsymbol{\gamma}}, \boldsymbol{u}, \boldsymbol{x}),$$

$$\langle F(\tilde{\boldsymbol{\gamma}}, \boldsymbol{u}, \boldsymbol{x})\rangle_u = \int d\boldsymbol{u} P(\boldsymbol{u}) F(\tilde{\boldsymbol{\gamma}}, \boldsymbol{u}, \boldsymbol{x}), \tag{41}$$

$$P(\tilde{\boldsymbol{\gamma}}) = \frac{e^{-\frac{1}{2}\tilde{\boldsymbol{\gamma}}\tilde{D}^{-1}\tilde{\boldsymbol{\gamma}}}}{\sqrt{(2\pi)^N det\tilde{D}}}, \tag{42}$$

$$P(\boldsymbol{u}) = \frac{e^{-\frac{1}{2}\mathbf{u}(\mathbf{B+bI})^{-1}\mathbf{u}}}{\sqrt{(2\pi)^N det(B+bI)}}. \tag{43}$$

As it is evident from eqs.(30),(40) the mutual information is expressed as series of Gaussian averages, where all powers higher than the second one can be treated via the Wick theorem. This allows to establish the basic rules for a diagrammatic expansion in terms of Feynman graphs, which is the subject of the next sections.

4   A diagrammatic formalism for the expansion

It is well known that higher order moments of a Gaussian distribution can be reduced to a series of products of second order moments via the use of Wick's theorem, both in classical and in quantum systems [12]. The introduction of a diagrammatic formalism allows to associate a graph to each type of integral. Therefore the whole series can be expressed in a very compact and elegant way and integrations can be performed symbolically contracting all the lines pair by pair, in such a way to obtain all the topologically distinct diagrams. Since each contraction is accociated to a precise numerical value, the value of each diagram can be easily calculated by simply multiplying all the factors corresponding to the contractions of the lines.

In the previous work [9], where we have focused on the first order terms in $\boldsymbol{g}$, in eqs.(30),(40), we have identified the basic elements which might allow to build a diagrammatic expansion for the mutual information, up to first order. Here we generalize our approach and we show that a diagrammatic interpretation can be provided for any order in the expansion, establishing the basic elements and rules distinctly for the output entropy and for the equivocation.

4.1   Diagrammatic rules for the evaluation of the output entropy

Since two distinct averages characterize eq.(30), namely on $\boldsymbol{\gamma}$ and $\boldsymbol{V}$, one is naturally tempted to introduce two distinct symbolic lines for $\boldsymbol{\gamma}$ and $\boldsymbol{V}$. Yet **two** distinct objects are to be contracted and integrated on $\boldsymbol{V}$, namely $\boldsymbol{V}$ itself and $D\boldsymbol{V}$. Therefore it is more convenient to introduce two distinct lines for $\boldsymbol{V}$ and $D\boldsymbol{V}$. Since either of this two lines can be contracted with itself or with the other one, one has three distinct rules on contraction corresponding to integrating different objects on $\boldsymbol{V}$. The whole prescription can be given as follows:

1. Each term V_i is represented by a straight line $\quad i \,\rule[0.5ex]{1.2em}{0.4pt}$
   $$V_i$$

2. Each term $(D\boldsymbol{V})_i$ is represented by a crossed solid line $\quad i \,\rule[0.5ex]{1em}{0.4pt}\!\times\!\rule[0.5ex]{1em}{0.4pt}$
   $$(D\boldsymbol{V})_i$$

3. Each term γ_i is represented by a wiggly line $\quad i \,\sim\!\sim\!\sim$
   $$\gamma_i$$
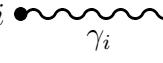
4. Each matrix element $[A+bI]^{-1}_{ij}$ is represented by a dashed romboid $\qquad i \,\diamondsuit\, j$

9

5. The integration of the product $V_i V_j$ over $\boldsymbol{V}$ corresponds to the contraction of two straight lines coming out of vertices i,j.

   This produces a matrix element $(A + bI)_{ij}$ $\quad i \,\bullet\!\!-\!\!\!-\!\!\!-\!\!\bullet\, j$

6. The integration of the product $V_i (D\boldsymbol{V})_j / b$ over $\boldsymbol{V}$ corresponds to the contraction of a straight line coming out of the vertex i with a straight crossed line coming out of vertex j.

   This produces a matrix element $[(A + bI)D]_{ij}/b$ $\quad i \,\bullet\!\!-\!\!\times\!\!-\!\!\bullet\, j$

7. The integration of the product $(D\boldsymbol{V})_i (D\boldsymbol{V})_j / b^2$ over $\boldsymbol{V}$ corresponds to the contraction of two straight crossed line coming out of the vertices i,j.

   This produces a matrix element $[D(A + bI)D^T]_{ij}/b^2$ $\quad i \,\bullet\!\!-\!\!\times\!\!-\!\!\times\!\!-\!\!\bullet\, j$

8. The integration of the product $\gamma_i \gamma_j$ over $\boldsymbol{\gamma}$ corresponds to the contraction of two wiggly lines coming out of vertices i,j.

   This produces a matrix element D_{ij} $\quad i \,\bullet\!\!\sim\!\!\sim\!\!\sim\!\!\bullet\, j$

Eq.(30) can be written in terms of these symbols:

$$
\begin{aligned}
H(\boldsymbol{V}) = H_0(\boldsymbol{V}) + \frac{1}{2} \Bigg\langle\!\Bigg\langle\!\Bigg\langle \sum_{l=1}^{\infty} \frac{g_0^l}{l!\,b^l} \sum_{ijk} &\Big[(\, i \,\bullet\!\!-\!\!-\, - \, i \,\bullet\!\!\sim\!\!\sim\, - \, i \,\bullet\!\!-\!\!\times\!\!-\,) \\
(\, i \,\bullet\!\!\sim\!\!\sim\, + \, i \,\bullet\!\!-\!\!\times\!\!-\,)^3 &- \frac{g_0}{2} (\, i \,\bullet\!\!\sim\!\!\sim\, + \, i \,\bullet\!\!-\!\!\times\!\!-\,)^6 \Big]^l \Bigg\rangle_{\!\!\bullet\sim} \Big(\, {}_j \!\!\diamondsuit\!\! {}_k \,\Big) \Bigg\rangle_{\!\!\bullet-} \Bigg\rangle_{\!\!\bullet\times} \\
+ \Bigg\langle\!\Bigg\langle \sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n(n-1)} &\Big\langle \sum_{l=1}^{\infty} \frac{g_0^l}{l!\,b^l} \sum_{i} \Big[(\, i \,\bullet\!\!-\!\!-\, - \, i \,\bullet\!\!\sim\!\!\sim\, - \, i \,\bullet\!\!-\!\!\times\!\!-\,) \\
(\, i \,\bullet\!\!\sim\!\!\sim\, + \, i \,\bullet\!\!-\!\!\times\!\!-\,)^3 &- \frac{g_0}{2} (\, i \,\bullet\!\!\sim\!\!\sim\, + \, i \,\bullet\!\!-\!\!\times\!\!-\,)^6 \Big]^l \Big\rangle^{n}_{\!\!\bullet\sim} \Bigg\rangle_{\!\!\bullet-} \Bigg\rangle_{\!\!\bullet\times},
\end{aligned}
\tag{44}
$$

where we have symbolically separated the average across $\boldsymbol{V}$ from the average across $D\boldsymbol{V}$ to remind that contractions have to be performed on both objects according to the rules given above.

4.2 Diagrammatic rules for the evaluation of the equivocation

A diagrammatic interpretation can be given for eq.(40) introducing proper symbols and rules for the contractions:

1. Each term u_i is represented by a double straight line $\quad i \,\bullet\!\!=\!\!=\!\!=$
   $$ u_i $$

2. Each term $(\tilde{D}\boldsymbol{u})_i / b$ is represented by a crossed double straight line $\quad i \,\bullet\!\!=\!\!=\!\!\times\!\!=\!\!=$
   $$ (\tilde{D}\boldsymbol{u})_i / b $$

3. Each term $\tilde{\gamma}_i$ is represented by a double wiggly line $\quad i \,\bullet\!\!\approx\!\!\approx\!\!\approx$
   $$ \tilde{\gamma}_i $$

4. Each term $(J\boldsymbol{x})_i$ is represented by a dotted line $\quad i \,\bullet \cdots\cdots\cdots$
   $$ (J\boldsymbol{x})_i $$

10

5. Each matrix element $[B + bI]^{-1}_{ij}$ is represented by a filled romboid i ◆ j

6. The integration of the product $u_i u_j$ over $\boldsymbol{u}$ corresponds to the contraction of two double straight lines coming out of vertices i,j.

   This produces a matrix element $(B + bI)_{ij}$ i ▬▬ j

7. The integration of the product $u_i(\tilde{D}\boldsymbol{u})_j/b$ over $\boldsymbol{u}$ corresponds to the contraction of a double straight line coming out of the vertex i with a double straight crossed line coming out of vertex j.

   This produces a matrix element B_{ij} i ▬✕▬ j

8. The integration of the product $(\tilde{D}\boldsymbol{u})_i(\tilde{D}\boldsymbol{u})_j$ over $\boldsymbol{u}$ corresponds to the contraction of two crossed double straight lines coming out of the vertices i,j.

   This produces a matrix element $[B(B + bI)^{-1}B^T]_{ij}$ i ▬✕✕▬ j

9. the integration of the product $\tilde{\gamma}_i\tilde{\gamma}_j$ over $\tilde{\boldsymbol{\gamma}}$ corresponds to the contraction of two double wiggly lines coming out of vertices i,j.

   This produces a matrix element $\tilde{D}_{ij}$ i ∿∿∿ j

10. The integration of the product $(J\boldsymbol{x})_i(J\boldsymbol{x})_j$ over $\boldsymbol{x}$ corresponds to the contraction of two dotted lines coming out of the vertices i,j.

    This produces a matrix element $[JCJ^T]_{ij}$ i •·········• j

Eq.(40) can be written in this formalism:

$$\langle H(\boldsymbol{V}|\boldsymbol{x})\rangle_x = \langle H_0(\boldsymbol{V}|\boldsymbol{x})\rangle_x + \frac{1}{2}\left\langle\left\langle\left\langle\left\langle\sum_{l=1}^{\infty}\frac{g_0^l}{l!\,b^l}\sum_{ijk}[(i\,\text{▬▬} - i\,\text{∿∿∿} - i\,\text{▬✕▬})\right.\right.\right.\right. \tag{45}$$

$$(i\,\text{∿∿} + i\,\text{▬✕▬} + i\,\text{•····})^3 - \frac{g_0}{2}(i\,\text{∿∿∿} + i\,\text{•✕▬} + i\,\text{•····})^6\bigg]^l\bigg\rangle_{\text{∿∿}}$$

$$\left(\text{▬◆▬}_{j\quad k}\right)\bigg\rangle_{\text{▬▬}}\bigg\rangle_{\text{•✕▬}}\bigg\rangle_{\text{•···}} + \left\langle\left\langle\left\langle\sum_{n=2}^{\infty}\frac{(-1)^{n+1}}{n(n-1)}\left\langle\sum_{l=1}^{\infty}\frac{g_0^l}{l!\,b^l}\sum_i[(i\,\text{▬▬} - i\,\text{∿∿∿} - i\,\text{▬✕▬})\right.\right.\right.\right.$$

$$(i\,\text{∿∿∿} + i\,\text{▬✕▬} + i\,\text{•····})^3 - \frac{g_0}{2}(i\,\text{∿∿} + i\,\text{▬✕▬} + i\,\text{•····})^6\bigg]^l\bigg\rangle_{\text{∿∿}}^n\bigg\rangle_{\text{▬▬}}\bigg\rangle_{\text{•✕▬}}\bigg\rangle_{\text{•···}}.$$

Eqs.(44),(45) constitute the final expression for the mutual information at every order. Application of the Wick's theorem and of the contraction rules allows to analytically derive the contribution to the mutual information at each order in $\boldsymbol{g}$.

The first order approximation in g_0, studied in [9], can be easily obtained from eqs.(45),(44), putting $l = 1$ and neglecting the double summation over the indices l and n, which gives corrections only at orders higher than first one in g_0.

5 A detailed analysis of the different contributions to the mutual information

The expansion we have derived allows to investigate how the different orders contribute to the mutual information, varying the expansion parameter g_0.

The expression of the mutual information at the zeroth order can be easily derived from eqs.(26),(37):

$$I_0 = H_0(\boldsymbol{V}) - \langle H_0(\boldsymbol{V}|\boldsymbol{x})\rangle_x = \frac{1}{2}\ln det\left(\frac{A+bI}{B+bI}\right);\tag{46}$$

The first order approximation has been a primary object of investigation in [9], where a diagrammatic interpretation has been provided, as well. From eqs.(27),(38) it can be shown that the final expression of the first order contribution to the mutual information can be written as follows:

$$I_1 = -3g_0 \sum_{ij} A_{ii}[[A+bI]_{ij}^{-1}A_{ij} - [B+bI]_{ij}^{-1}B_{ij}],\tag{47}$$

Further details about the derivation of the first order approximation can be found in [9].

We now focus on the second order contributions. Let us reconsider eqs.(44) and (45). It is clear that the expansion in l, n is not a direct expansion in g_0: the first order in l contains both first and second order terms in g_0. Therefore one must be careful and extract all the second order terms in g_0 from the proper powers in l, n. In particular, in the expansion for both, the output entropy and the equivocation, one has to put $l = 1, 2$ in the first sum and $n = 2, l = 1$ in the second sum, and then retain only the second order terms. It can be shown that:

$$
H_2(\boldsymbol{V}) = \frac{g_0^2}{b^2}\left\{\frac{1}{4}\left\langle\left\langle\left\langle\sum_{ijk}(i\,\text{●——}\,-\,i\,\text{●∿∿}\,-\,i\,\text{●—✳—})^2(i\,\text{●∿∿}\,+\,i\,\text{●—✳—})^6\right.\right.\right.\right.
$$
$$
-b\,(i\,\text{●∿∿}\,+\,i\,\text{●—✳—})^6\Big\rangle_{\!\text{●∿}}\Big(\underset{j}{\overline{}}\text{◈}\underset{k}{\overline{}}\Big)\Big\rangle_{\!\text{●—}}\Big\rangle_{\!\text{●✳}}
$$
$$
-\frac{1}{2}\left\langle\left\langle\left\langle\sum_i(i\,\text{●——}\,-\,i\,\text{●∿∿}\,-\,i\,\text{●—✳—})(i\,\text{●∿∿}\,+\,i\,\text{●—✳—})^3\right\rangle^2_{\!\text{●∿}}\right\rangle_{\!\text{●—}}\right\rangle_{\!\text{●✳}}\right\}\tag{48}
$$

$$\langle H_2(\boldsymbol{V}|\boldsymbol{x})\rangle_x =$$
$$
\frac{g_0^2}{b^2}\left\{\frac{1}{4}\left\langle\left\langle\left\langle\left\langle\sum_{ijk}(i\,\text{●══}\,-\,i\,\text{●∿∿}\,-\,i\,\text{●—✳═})^2(i\,\text{●∿∿}\,+\,i\,\text{●—✳═}\,+\,i\,\text{●·····})^6\right.\right.\right.\right.\right.
$$
$$
-b\,(i\,\text{●∿∿}\,+\,i\,\text{●—✳═}\,+\,i\,\text{●·····})^6\Big\rangle_{\!\text{●∿}}\Big(\underset{j}{\overline{}}\text{◆}\underset{k}{\overline{}}\Big)\Big\rangle_{\!\text{●—}}\Big\rangle_{\!\text{●✳}}\Big\rangle_{\!\text{●...}}
$$
$$
-\frac{1}{2}\left\langle\left\langle\left\langle\left\langle\sum_i(i\,\text{●══}\,-\,i\,\text{●∿∿}\,-\,i\,\text{●—✳═})\right.\right.\right.\right.
$$
$$
(i\,\text{●∿∿}\,+\,i\,\text{●—✳═}\,+\,i\,\text{●·····})^3\Big\rangle_{\!\text{●∿}}\Big\rangle_{\!\text{●—}}\Big\rangle_{\!\text{●✳}}\Big\rangle_{\!\text{●...}}\right\},\tag{49}
$$

where we have extracted the terms which are order g_0^2.

The computation of all the second order diagrams is very long, since it involves sixth powers of terms and 4 different types of contractions, both in the equivocation and in the output entropy. Yet the calculation results much simplified in some limits.

In particular, let us consider the case where $g_0 \ll b$ but b is not too small. Both the equivocation and the output entropy contain terms of order g_0^2/b^2 up to $g_0^2 b^2$. In fact, looking at the contraction rules given above, one can notice that the contraction of two wiggly lines produces a matrix element D_{ij}, where the matrix D is order b, while the other contractions produce elements which are order 1 with respect to b. Therefore the dominating contributions in the limit when b is not too small are given by the diagrams where 3 or 4 wiggly loops appear. Keeping only these diagrams, after having performed all the contractions, it can be shown that the dominating second order contribution to the output entropy is given by the following sum of diagrams:

$$H_2(\boldsymbol{V}) \simeq -\frac{9}{2}\frac{g_0^2}{b^2}\left[2\sum_{ij} \begin{matrix} i \\ \end{matrix}\begin{matrix} \\ j \end{matrix}\right], \tag{50}$$

which corresponds to the following analytical expression:

$$H_2(\boldsymbol{V}) \simeq -\frac{9}{2}g_0^2 b^2 \left[Tr\left(A\left(A+bI\right)^{-1}\right)^2\right]^2. \tag{51}$$

Under the same assumption the second order contribution to the equivocation can also be expressed as a simple sum of diagrams:

$$\langle H(\boldsymbol{V}|\boldsymbol{x})\rangle_x \simeq -\frac{g_0^2}{b^2}\sum_i \left[\frac{15b}{4}\ \begin{matrix} i \end{matrix}\ + 7\ \begin{matrix} i \end{matrix}\right]. \tag{52}$$

After some elementary manipulation of the matrices it is easy to show that the analytical expression for the equivocation reads:

$$\langle H(\boldsymbol{V}|\boldsymbol{x})\rangle_x \simeq -g_0^2 b^2 \left[\frac{15}{4}Tr\left(B\left(B+bI\right)^{-1}\right)^3 + 7\,Tr\left(B\left(B+bI\right)^{-1}\right)^4\right]; \tag{53}$$

As a whole, the second order contribution to the mutual information in the limit when b is not too small can be written:

$$I_2 \simeq 3g_0^2 b^2 \left(\frac{5}{4}Tr\left(B\left(B+bI\right)^{-1}\right)^3 + \frac{7}{3}Tr\left(B\left(B+bI\right)^{-1}\right)^4 - \frac{3}{2}\left[Tr\left(A\left(A+bI\right)^{-1}\right)^2\right]^2\right); \tag{54}$$

Since the evaluation of the mutual information has been carried out for a generic connectivity matrix $\{J_{ij}\}$ it is obvious that both the total information value and its order-specific

13
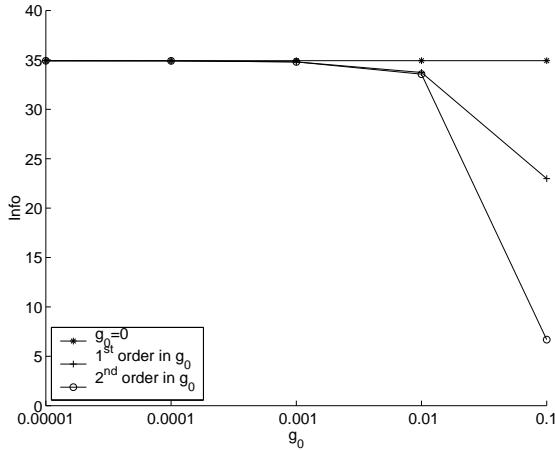
Figure 1: Mutual information as a function of the parameter g_0, comparing the zeroth, first and second order approximation. $N = 100; b_0 = 0.1; b = 0.6$. The correlation matrix C is equal to the identity and the connectivities J_{ij} linking each output unit i to the input neurons j are randomly chosen from a Gaussian distribution with mean zero and standard deviation $1/\sqrt{N}$

contributions will depend on the structure of the connectivities. Even without aiming at a generalization we can in any case provide a quantification of the different contributions to the mutual information restricting ourselves to the specific case where the connectivities linking each output neuron to the input ones are drawn from a gaussian distribution with mean zero and standard deviation $1/\sqrt{N}$. As it is known from the theory of spin glasses and neural networks, this renormalization ensures that the local field h_i acting on each unit i is finite when N is large.

Fig.(1) shows the mutual information in zeroth, first and second order approximation for increasing values of g_0. Two main observations stem from the analysis of the curves:

- both the first and the second order contributions in the non-linearity lower the information value with respect to the linear network.

- our approximations start to lose their validity for values of g_0 larger than 0.01; this does not automatically mean that one should add higher order contributions, since we have neglected second order contributions with powers of b lower than b^2: when g_0 becomes not much smaller than b one should probably include the other contributions, at second order in g_0.

Fig.(2) shows a detail of the previous plot, where the linear and quadratic fit in g_0 are more evident.

Fig.(3) shows the mutual information as a function of the population size N. As it is clear from eqs.(46),(47),(54), the zeroth order approximation is linear in N, and so is also the first order contribution, since it is the difference between two scalar products of N-dimensional vectors. On the other hand the second order contribution is roughly quadratic in N: in fact a numerical check of the three different contributions shows that the first two terms, which

14
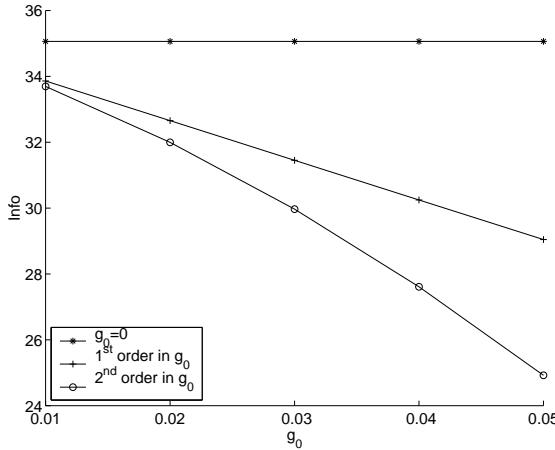
Figure 2: Mutual information as a function of the parameter g_0, as in fig.(1). $N = 100; b_0 = 0.1; b = 0.6$. The correlation matrix C is equal to the identity and the connectivities J_{ij} linking each output unit i to the input neurons j are randomly chosen from a Gaussian distribution with mean zero and standard deviation $1/\sqrt{N}$

are linear in N (traces of N-dim matrices) are three orders of magnitudes smaller than the third term, which is quadratic in N (square of a trace of an N-dim matrix).

6 Conclusions

We have presented here a systematic approach to quantify the effect of small non-linearities in the transfer function on the information transferred by a two layer network of analogue units. We have derived a perturbative expansion in the non-linearity parameter g_0, providing an elegant interpretation in terms of Feynman diagrams.

In a previous report [9] we had already quantified the contribution to the information at first order in g_0. Here we have extended the previous results providing an analytical expression to calculate the contributions at each order in perturbation theory. Moreover our method can be easily applied to any structure of the connectivities, with no restriction to a specific architecture.

As an example, we have quantified the zeroth, first and second order contributions to the information in the case of random, Gaussian distributed couplings and in the limit when the output noise b is not small. We have found that the main effect of the non-linearity for this particular architecture is a loss in information, detected both at first and at second order in g_0. This result is in agreement with previous investigations [3], where it has been shown that two main causes of information loss in a two-layer network with random gaussian couplings are a non-linearity in the transfer function, like the presence of a threshold, and a large output noise.

The detailed analysis we have presented here applies to the particular case of cubic non-linearities. Yet, as already remarked in [9], our method can be easily adapted to any non-linearity of a generic power $2k + 1$: it is enough to replace the third and sixth powers appearing in eqs.(44) and (45) respectively with powers $2k + 1$ and $4k + 2$.

15
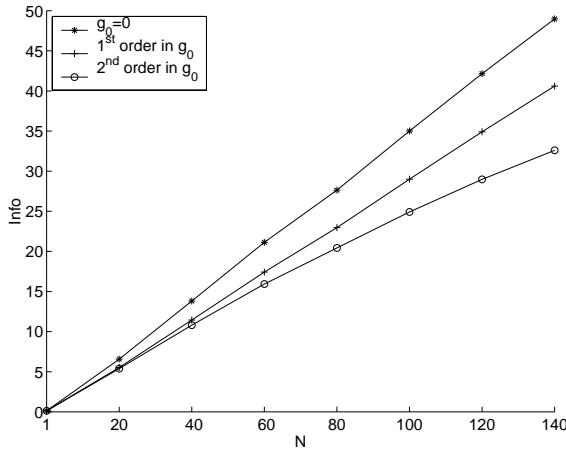
Figure 3: Mutual information as a function of the population size N, comparing the zeroth, first and second order approximations in g_0. $g_0 = 0.05; b_0 = 0.1; b = 0.6$. The correlation matrix C is equal to the identity and the connectivities J_{ij} linking each output unit i to the input neurons j are randomly chosen from a Gaussian distribution with mean zero and standard deviation $1/\sqrt{N}$

Finally this allows to deal with highly non-linear functions, like a hyperbolic tangent, which has often been proposed in modelling realistic neural systems [10].

Acknowledgments

References

[1] S. Schultz and A. Treves. Stability of the replica symmetric solution for the information conveyed by a neural network. *Physical Review E*, 57(3):3302–3310, 1998.

[2] A. Treves. Quantitative estimate of the information relayed by the schaffer collaterals. *J. Comp. Neurosci.*, 2:259–272, 1995.

[3] V. Del Prete and A. Treves. Replica symmetric evaluation of the information transfer in a two-layer network in presence of continuous and discrete stimuli. *Phys. Rev. E*, 2002. in press.

[4] A Campa, P. Del Giudice, N. Parga, and J.P. Nadal. *Network*, 6:449, 1995.

[5] J. H. van Herten. *J. Comp. Physiology*, A171:157, 1992.

[6] J. P. Nadal and N. Parga. *Network*, 4(3):295, 1993.

[7] E. Korutcheva, J. P. Nadal, and N. Parga. *Network*, 8:405, 1997.

[8] A. Turiel, E. Korutcheva, and N. Parga. *J. Phys.A: Math. Gen.*, 32:1875, 1999.

[9] E. Korutcheva, V. Del Prete, and J.P. Nadal. A perturbative approach to non-linearities in the information carried by a two layer neural network. *International Journal of Modern Physics B*, 15(3):281–295, 2001.

[10] Daniel J. Amit. *Modeling Brain Function.* Cambridge University Press, Cambridge, 1989.

[11] J. A. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the Theory of Neural Computation.* Addison-Wesley Publishing Company, Santa Fe Institute, 1991. Lecture Notes vol.1.

[12] A. Abrikosov, L. Gor'kov, and I. Dzyaloshinskij. *Quantum field theoretical methods in statistical physics, 2nd ed.* Pergamon Press, Oxford, 1965.